On a Multi–Objective Evolutionary Algorithm and Its Convergence to the Pareto Set

Günter Rudolph

Abstract—**Although there are many versions of evolutionary algorithms that are tailored to multi–criteria optimization, theoretical results are apparently not yet available. Here, it is shown that results known from the theory of evolutionary algorithms in case of single criterion optimization do not carry over to the multi–criterion case. At first, three different step size rules are investigated numerically for a selected problem with two conflicting objectives. The empirical results obtained by these experiments lead to the observation that only one of these step size rules may have the property to ensure convergence to the Pareto set. A theoretical analysis finally shows that a special version of an evolutionary algorithm with this step size rule converges with probability one to the Pareto set for the test problem under consideration.**

Keywords—**multi–criteria optimization, evolutionary algorithms, stochastic convergence to Pareto set**

I. Introduction

In classical optimization theory it is tacitly assumed that there exists a single global objective function. But in practical problems the decision maker is almost always faced with multiple—usually conflicting—goals. In this situation one often aggregates the vector–valued objective function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ $(m > 1)$ into a scalar–valued surrogate objective function $f : \mathbb{R}^n \to \mathbb{R}$, for example, via $f(\mathbf{x}) = \mathbf{w}' \mathbf{f}(\mathbf{x})$ with some weight vector $\mathbf{w} \in \mathbb{R}^m$. Although this approach opens the door for single criteria optimization methods on the one hand, it also introduces a not negligible degree of uncertainty for the decision maker on the other hand: One cannot be sure whether the chosen weights do reflect the importance of each original goal appropriately. In fact, as soon as a specific weight vector has been chosen, the original decision space is considerably and prematurely cut down *before* enough information could be gathered that might justify such a reduction. As a consequence, a huge number of potential good decisions are precluded *a priori*.

Another approach to attack the problem takes into account that the vectors of objective function values $\mathbf{y} = \mathbf{f}(\mathbf{x})$ are partially ordered. An objective vector $\mathbf{y}$ is said to dominate objective vector $\tilde{\mathbf{y}}$ if $y_i \le \tilde{y}_i$ for all $i = 1, \ldots, m$ and $y_i < \tilde{y}_i$ for at least one index i. A decision vector $\mathbf{x}$ with $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is called Pareto–optimal if there is no decision vector $\tilde{\mathbf{x}} \in \mathbb{R}^n$ for which $\tilde{\mathbf{y}} = \mathbf{f}(\tilde{\mathbf{x}})$ dominates $\mathbf{y}$. The set of all Pareto–optimal decision vectors is termed the Pareto–optimal, efficient, or admissible set of the problem. The corresponding set of objective vectors is called the nondominated set.

If it is possible to determine the Pareto–optimal set (or short: Pareto set), then the decision maker has the opportunity to learn about the tradeoffs being associated with the problem at hand so that a ranking of the decision maker's true preferences can be made *after* all possible solutions are known. Although the Pareto set may be determined analytically in exceptional cases, one has to seek remedy in numerical approximation methods in general. It has been demonstrated several times that some versions of evolutionary algorithms are able to accomplish this task to a reasonable degree. As can be learned from the recent surveys in [1], [2], [3] there are numerous suggestions of multi–objective EAs. Moreover, there are some vague empirical rules indicating which version works better than some other under certain circumstances whereas theoretical results are apparently rare. For example, it can be shown that an EA generates at least one stochastic trajectory converging to the Pareto set with probability one if the search space is finite, the support of the fixed mutation distribution covers the search space, and offsprings are accepted if they dominate the parents [4]. Since these assumption are rather strong it may be instructive to investigate what happens if some of these assumptions are weakened.

Here, the analysis will focus on a *simplified version* of a multi–objective EA originally presented in [5]. Assume there are two objective functions f_0 and f_1, both to be minimized. Let $\mathbf{X}_k \in \mathbb{R}^n$ be some individual at generation $k \ge 0$. An offspring $\mathbf{Y}_k$ is generated as follows: After having drawn a random vector $\mathbf{M}$ with zero mean and a random variable J with $\mathsf{P}\{ J = 0 \} = \mathsf{P}\{ J = 1 \} = 1/2$ one sets

$$\mathbf{Y}_k = \mathbf{X}_k + s(J, \mathbf{X}_k) \cdot \mathbf{M}$$

where the distribution of $\mathbf{M}$ is fixed over time and the step size $s > 0$ may depend on J and $\mathbf{X}_k$. If $f_J(\mathbf{Y}_k) < f_J(\mathbf{X}_k)$ then the offspring $\mathbf{Y}_k$ is accepted, otherwise rejected. Then the process repeats.

Evidently, this method may be seen as a multi–objective generalization of the $(1+1)$–EA as it is known in single criterion optimization. One might be tempted to speculate that the existing theory for the single criterion $(1+1)$–EA is easily transferable to the multiple criteria version. But this is a misspeculation. For example, it is numerically shown in Section II that the step size control must be changed significantly. Moreover, the old notion of "convergence to the optimum" needs a reformulation before a theoretical convergence analysis, given in Section III, can begin.

II. Numerical Preliminary Study

Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ be a vector–valued function with

$$\mathbf{f}(\mathbf{x}) = (\, \|\mathbf{x}\|^2, \|\mathbf{x} - \mathbf{z}\|^2 \,)' \tag{1}$$

and $\mathbf{0} \ne \mathbf{z} \in \mathbb{R}^2$. Assume that the two objective functions are to be minimized. The efficient set can be determined analytically and it is given by

$$\mathcal{X}^* = \{\, \mathbf{x} \in \mathbb{R}^2 : \mathbf{x} = r\,\mathbf{z}, r \in [\,0, 1\,] \,\}\,.$$

In the sequel three different step size rules and two mutation distributions are compared numerically. The first mutation distribution is the two–dimensional Gaussian distribution with zero mean and the unit matrix as covariance matrix. The second mutation distribution is the uniform distribution on the unit circle.

G. Rudolph is with the Department of Computer Science, University of Dortmund, Germany. E-mail: rudolph@icd.de . This work is a result of the *Collaborative Research Center SFB 531* at the University of Dortmund. Financial support by the German National Science Foundation (DFG) is gratefully acknowledged.

To distinguish between the different distributions of the mutation vector we shall write **G** resp. **U** in lieu of **M**. A two–dimensional random vector $\mathbf{G} = (G_1, G_2)'$ is generated by drawing two independent standard normal random variables G_1 and G_2, whereas random vector $\mathbf{U} = (U_1, U_2)'$ is obtained by drawing random variable ω uniformly distributed in $[\,0, 2\,\pi)$ and setting $U_1 = \cos\omega$ and $U_2 = \sin\omega$. Since both distributions are spherically symmetric one may choose $\mathbf{z} = (z_1, 0)'$ with $z_1 > 0$ without affecting the generality of the results—but the theoretical analysis in the next section will be considerably simplified. For example, the distance $d(\mathbf{x}, \mathcal{X}^*) = \min\{||\mathbf{x}-\mathbf{x}^*|| : \mathbf{x}^* \in \mathcal{X}^*\}$ of some point **x** to the Pareto set $\mathcal{X}^*$ reduces to

$$d(\mathbf{x}, \mathcal{X}^*) = \begin{cases} ||\mathbf{x}|| & \text{if } x_1 < 0 \\ |x_2| & \text{if } 0 \le x_1 \le z_1 \\ ||\mathbf{x}-\mathbf{z}|| & \text{if } x_1 > z_1\,. \end{cases} \qquad (2)$$

For the sake of brevity, the efficient set will be omitted from the argument list, i.e., we shall simply write $d(\mathbf{x})$ to denote the distance to the efficient set.

A. *Fixed Step Size*

In single criterion optimization it is known that the $(1{+}1)$–EA with fixed variance $\sigma^2 > 0$ and Gaussian mutations converges with probability one to the global optimum under weak regularity assumptions. This property does not hold for fixed step size $s > 0$ with uniform mutations on the surface of a unit hyperball. This difference remains valid even for convex objective functions. One might speculate that these result carry over in some manner to multi–objective EAs. Two simple numerical experiments, however, reveal that this hypothesis must be rejected.

If $\mathbf{z} = (100, 0)'$ then the efficient set is located on the x_1–axis between 0 and 100. The initial point is set to $\mathbf{X}_0 = (300, 300)'$, the step size is set to $s = 10$ and the variance σ^2 is chosen such that $\mathsf{E}[\,||\sigma\,\mathbf{G}||\,] = \sigma\,\sqrt{\pi/2}$ equals $\mathsf{E}[\,||s\,\mathbf{U}||\,] = s$. The upper graphs of Figs. 1 & 2 show a typical run in case of Gaussian and uniform mutations, respectively. One might be tempted to deduce from these curves that the $(1+1)$–EA has got stuck prematurely in some small region. But this is not the case. Rather, the sequence $(\mathbf{X}_k : k \ge 0)$ cycles between the two single criterion optima $(0, 0)'$ and $(100, 0)'$ as it may become visible from the lower graphs of Figs. 1 & 2. In any case, there is no convergence to the Pareto set regardless which mutation distribution (Gaussian or uniform) is chosen.

B. *Optimal Single Objective Step Sizes*

Since both objective functions are strongly convex it might appear reasonable to exploit the existing theory for the single criterion $(1+1)$–EA. Following [6, p. 170f.] the optimal step size for uniform mutations is

$$s^*(\mathbf{X}_k, J) = \frac{c^*}{2}\,||\nabla f_J(\mathbf{X}_k)|| \approx 0.39424 \cdot ||\nabla f_J(\mathbf{X}_k)||$$

where $c^* = 0.78847$ with $g(c^*) = 0.7693$ is the minimizing solution of

$$g(c) = 1 - \frac{c\,\sqrt{4-c^2} + c^2\,\arcsin(c/2)}{\pi} + \frac{c^2}{2}\,. \qquad (3)$$

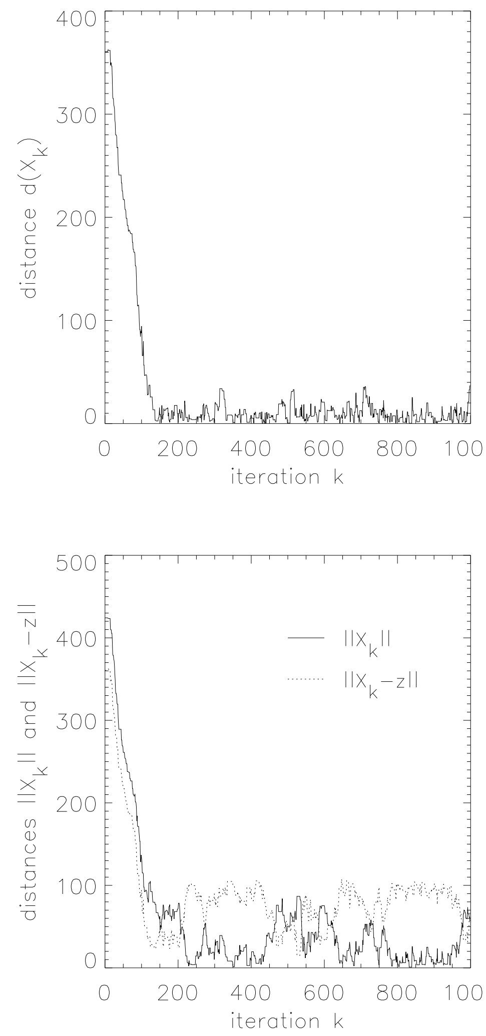

Fig. 1. Typical run of the multi–objective $(1+1)$–EA under Gaussian mutations with fixed variance. The upper graph shows the distances $d(\mathbf{X}_k)$ to the Pareto set during the run, while the lower one shows the distances $||\mathbf{X}_k||$ and $||\mathbf{X}_k - \mathbf{z}||$ to the single criterion solutions of f_0 and f_1.

In general, for arbitrary $c \in (0, 2)$ one obtains

$$\mathsf{E}[\,f_j(\mathbf{X}_{k+1})\,|\,X_k\,] = g(c)\,f_j(\mathbf{X}_k) \qquad (4)$$

which implies a geometrically fast approach to the optimum of the single criterion problem. In case of Gaussian mutations [6, p. 185f.] the relation

$$\mathsf{E}[\,f_j(\mathbf{X}_{k+1})\,|\,X_k\,] = 0.81185 \cdot f_j(\mathbf{X}_k)$$

with $\sigma_k^* \approx 0.26885 \cdot ||\nabla f_j(\mathbf{X}_k)||$ is valid.

When using this type of step size or variance control for the multicriteria $(1+1)$–EA there is again hardly a difference in the convergence behavior between Gaussian and uniform mutations. Since the steps sizes (or variances) in the vicinity of the Pareto set are often much larger than $s = 10$, the sequence of parents $(\mathbf{X}_k : k \ge 0)$ now cycles more frequently between the minimizing points of the single criterion problems during 1000 iterations than the fixed step size version. Therefore, it is refrained from presenting plots of these runs. But it should be

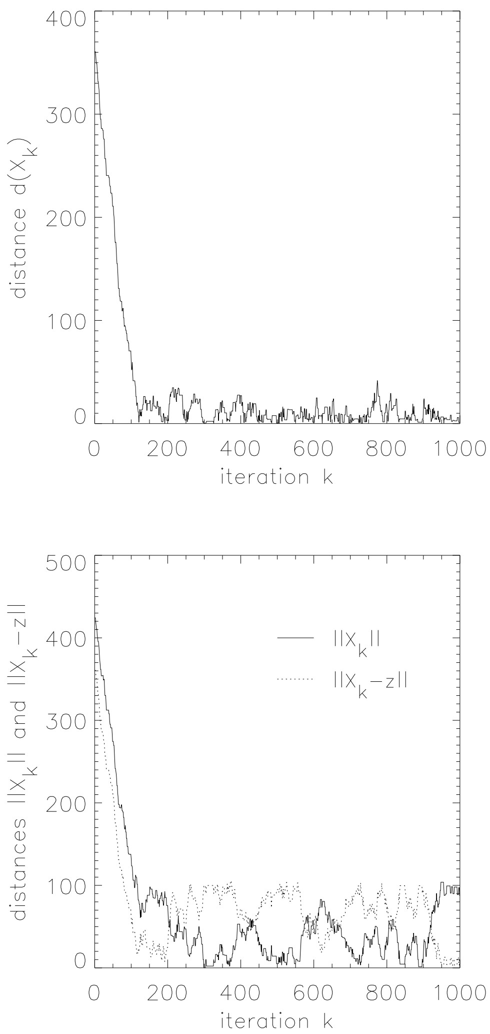

Fig. 2. Typical run of the multi–objective $(1+1)$–EA under uniform mutations with fixed step size. The upper graph shows the distances $d(\mathbf{X}_k)$ to the Pareto set during the run, while the lower one shows the distances $||\mathbf{X}_k||$ and $||\mathbf{X}_k - \mathbf{z}||$ to the single criterion solutions of f_0 and f_1.

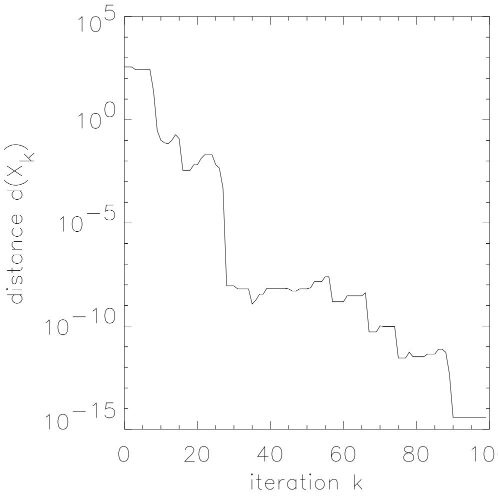

Fig. 3. The distance to the Pareto set of the multi–objective $(1 + 1)$–EA with step size rule $s = d(\mathbf{X}_k)$ under uniform mutations (typical run).

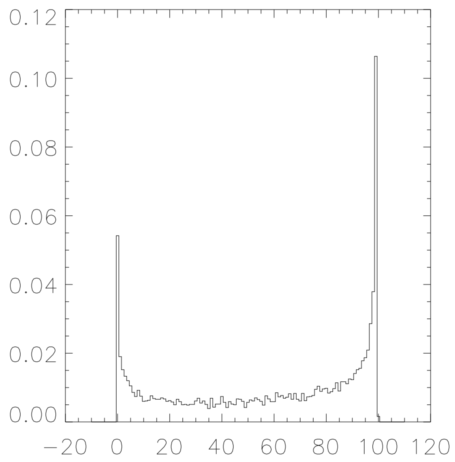

Fig. 4. The frequency distribution of the first component's value of the process stopped as soon as $d(\mathbf{X}_k) < 10^{-8}$.

noted that the average distance to the Pareto set is much larger than using fixed step sizes. As a consequence, there is no convergence to the Pareto set when using this type of step size rules.

C. Step Sizes Proportional to Distance to Pareto Set

A closer look at the upper graphs of Figs. 1 & 2 reveals that the average distance to the Pareto set is approximately of the order of the fixed step size $s = 10$ as soon as the sequence $(\mathbf{X}_k)$ starts cycling. One might speculate that a decreasing of the step size would also decrease the average distance to the Pareto set. Therefore, the third step size rule is chosen as follows: $s_k = d(\mathbf{X}_k)$. As can be seen from Fig. 3, this step size rule offers a rapid approach to the Pareto set. Another noteworthy property of this step size rule is illustrated by the following experiment: The multicriteria $(1 + 1)$–EA is started from $\mathbf{X}_0 = (300, 300)'$ and it is stopped as soon as $d(\mathbf{X}_k) \leq \varepsilon = 10^{-8}$. Then the first component of vector $\mathbf{X}_k$ is stored to some file and the $(1+1)$–EA is restarted again. Fig. 4 summarizes the results obtained after 10,000 experiments. Evidently, the $(1 + 1)$–EA converges more frequently in the vicinity of the solutions of the single criterion problems, but it also converges to other points of the Pareto set. The most frequent realizations are close to **z**, which is closest to the starting point.

Summing up: After this numerical preliminary study one may conclude that only the third step size rule deserves a theoretical investigation. This is done in the next section.

III. Analysis

In the sequel the convergence analysis will be restricted to uniform mutations on the circle with step size $s = d(\mathbf{x}_k)$. But at first it must be made rigorous what is meant by the notion of "convergence to the Pareto set."

Definition 1: Let $(\mathbf{X}_k : k \geq 0)$ be the sequence of points generated by the multi–objective $(1 + 1)$–EA. The EA is said to converge (with probability 1, in probability, in mean, etc.) to the Pareto set if the random sequence $(D_k : k \geq 0)$ with $D_k = d(\mathbf{X}_k)$ converges (with probability 1, in probability, in mean, etc.) to zero. □

A convenient avenue to establish such a property is shown by a customized version of a convergence result proven in [7, pp. 83–84].

Theorem 1: Let $(D_k : k \geq 0)$ be a sequence of nonnegative random variables and let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function vanishing only at the origin. If $\mathsf{E}[\, D_k \,] < \infty$ and

$$\mathsf{E}[\, D_{k+1} \,|\, \mathcal{F}_k \,] \leq D_k - \gamma(D_k)$$

for all $k \geq 0$ then the sequence $(D_k : k \geq 0)$ converges to zero with probability one as $k \to \infty$. □

Thus, it is sufficient to find a function $\gamma(\cdot)$ such that

$$\mathsf{E}[\, d(\mathbf{X}_{k+1}) \,|\, \mathbf{X}_k = \mathbf{x} \,] \leq d(\mathbf{x}) - \gamma(d(\mathbf{x}))$$

for all $\mathbf{x} \in \mathbb{R}^2$. Notice that the distribution of $d(\mathbf{X}_{k+1})$ depends on the location of the current parent $\mathbf{x}_k$ and on the two random variables J and ω. The stochastic effect of J can be "eliminated" by further conditioning:

$$\begin{aligned} \mathsf{E}[\, D_{k+1} \,|\, \mathbf{X}_k = \mathbf{x} \,] \; = & \\ & \mathsf{E}[\, D_{k+1} \,|\, \mathbf{X}_k = \mathbf{x}, J = 0 \,] \times \mathsf{P}\{\, J = 0 \,\} \; + \\ & \mathsf{E}[\, D_{k+1} \,|\, \mathbf{X}_k = \mathbf{x}, J = 1 \,] \times \mathsf{P}\{\, J = 1 \,\}\,. \end{aligned} \tag{5}$$

But the analysis can be further simplified by exploiting the symmetries of our particular test problem. Since for all $\mathbf{x} \in \mathbb{R}^2$ and $z_1 > 0$ holds

$$\mathsf{E}[\, D_{t+1} \,|\, \mathbf{X}_t = (x_1, x_2), J = 1 \,] =$$
$$\mathsf{E}[\, D_{t+1} \,|\, \mathbf{X}_t = (z_1 - x_1, x_2), J = 0 \,] \tag{6}$$

the analysis may be restricted to selection with respect to objective function f_0. Moreover, it suffices to consider the case $x_2 \geq 0$ since for all $\mathbf{x} \in \mathbb{R}^2$ holds

$$\mathsf{E}[\, D_{t+1} \,|\, \mathbf{X}_t = (x_1, x_2) \,] = \mathsf{E}[\, D_{t+1} \,|\, \mathbf{X}_t = (x_1, -x_2) \,]\,.$$

As a consequence, one only needs to determine

$$\mathsf{E}[\, D_{k+1} \,|\, \mathbf{X}_k = \mathbf{x}, J = 0 \,] = \mathsf{E}_0[\, D_{k+1} \,]$$

for all $\mathbf{x}$ with $x_1 \in \mathbb{R}$ and $x_2 > 0$. The shorthand expression on the r.h.s. of the equation above will be used for notational convenience.

The analysis has to be split in several subcases. At first consider those $\mathbf{x} \in \mathbb{R}^2$ with $\mathbf{x} \in \mathcal{L} = \{\mathbf{y} \in \mathbb{R}^2 : y_1 < 0, y_2 > 0\}$. Their distances to the Pareto set are given by $d(\mathbf{x}) = ||\mathbf{x}||$. Since the step size is set to $s = d(\mathbf{x}) = ||\nabla f_0(\mathbf{x})||/2$ it follows from eqns. (4) and (3) that $\mathsf{E}[\, f_0(\mathbf{X}_{k+1}) \,] \leq g(1)\, f_0(\mathbf{x}) = g(1)\, d^2(\mathbf{x})$ with $g(1) \approx 0.7820$. Notice that the relation $0 \leq d^2(\mathbf{X}_{k+1}) \leq f_0(\mathbf{X}_{k+1})$ is valid in general. This leads to

$$\mathsf{E}_0[\, d^2(\mathbf{X}_{k+1}) \,] \leq \mathsf{E}[\, f_0(\mathbf{X}_{k+1}) \,] \leq g(1)\, d^2(\mathbf{x})\,. \tag{7}$$

Owing to Jensen's inequality for conditional expectations and taking the square root in (7) one finally obtains

$$\mathsf{E}_0[\, d(\mathbf{X}_{k+1}) \,] \leq \mathsf{E}_0[\, d^2(\mathbf{X}_{k+1}) \,]^{1/2} \leq \sqrt{g(1)}\, d(\mathbf{x})\,.$$

Now suppose that it can be shown that $\mathsf{E}_0[\, d(\mathbf{X}_{k+1}) \,] \leq d(\mathbf{x})$ for all $\mathbf{x}$ with $\mathbf{x} \in \mathcal{R} = \{\mathbf{y} \in \mathbb{R}^2 : x_1 > z_1, x_2 > 0\}$, i.e., $d(\mathbf{x}) = ||\mathbf{x} - \mathbf{z}||$. Owing to the symmetry property (6) this would imply $\mathsf{E}_1[\, d(\mathbf{X}_{k+1}) \,] \leq d(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{L}$. Insertion into (5) would yield

$$\begin{aligned} \mathsf{E}[\, D_{k+1} \,|\, \mathbf{X}_k = \mathbf{x} \,] &\leq \sqrt{g(1)}\, d(\mathbf{x}) \times \frac{1}{2} + d(\mathbf{x}) \times \frac{1}{2} \\ &= \frac{\sqrt{g(1)} + 1}{2} \cdot d(\mathbf{x}) \\ &< \frac{19}{20} \cdot d(\mathbf{x}) \end{aligned} \tag{8}$$

for all $\mathbf{x} \in \mathcal{L}$. But if inequality (8) holds for all $\mathbf{x} \in \mathcal{L}$ then it must also hold for all $\mathbf{x} \in \mathcal{R}$ which follows immediately from the symmetry property (6).

Thus, to verify inequality (8) it is necessary to prove inequality $\mathsf{E}_0[\, d(\mathbf{X}_{k+1}) \,] \leq d(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{R}$. Therefore let $\mathbf{x} \in \mathcal{R}$ so that the step size is set to $s = d(\mathbf{x}) = ||\mathbf{x} - \mathbf{z}||$. Since

$$X_{k+1} = \begin{cases} \mathbf{x} + s\,\mathbf{U} & \text{if } f_0(\mathbf{x} + s\,\mathbf{U}) < f_0(\mathbf{x}) \\ \mathbf{x} & \text{otherwise} \end{cases}$$

one obtains

$$d(X_{k+1}) = \begin{cases} d(\mathbf{x} + s\,\mathbf{U}) & \text{if } f_0(\mathbf{x} + s\,\mathbf{U}) < f_0(\mathbf{x}) \\ ||\mathbf{x} - \mathbf{z}|| & \text{otherwise} \end{cases}$$

for the random distance of the new parent provided selection is with respect to f_0. Recall from the previous section that $\mathbf{U} = (\cos\omega, \sin\omega)'$ where ω is uniformly distributed on $[\,0, 2\,\pi)$ and assume that (ω_0, ω_1) with $\omega_0 \leq \omega_1$ represents the interval with the property:

$$\omega \in (\omega_0, \omega_1) \iff f_0(\mathbf{x} + s\,\mathbf{U}) < f_0(\mathbf{x})\,.$$

Then the expectation may be obtained via

$$\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U}) \,] = \int_{\omega_0}^{\omega_1} \frac{d(\mathbf{x} + s\,\mathbf{U})}{2\,\pi}\, d\omega + ||\mathbf{x} - \mathbf{z}|| \left(1 - \frac{\omega_1 - \omega_0}{2\,\pi}\right).$$

Let $\mathbf{p}$ and $\mathbf{q}$ with $p_1 < q_1$ are the two points of intersection of the two circles $\{\mathbf{y} \in \mathbb{R}^2 : ||\mathbf{y} - \mathbf{x}|| = s\}$ and $\{\mathbf{y} \in \mathbb{R}^2 : ||\mathbf{y} - \mathbf{z}|| = ||\mathbf{x} - \mathbf{z}||\}$. Then the first angle is $\omega_0 = \arccos(h_0)$ with $h_0 = (p_1 - x_1)/||\mathbf{x} - \mathbf{z}||$ if $p_2 \geq x_2$ and $\omega_0 = \arcsin(h_0)$ otherwise, whereas the second angle is $\omega_1 = \arccos(h_1)$ with $h_1 = (q_1 - x_1)/||\mathbf{x} - \mathbf{z}||$ if $q_2 \geq x_2$ and $\omega_1 = \arcsin(h_1)$ otherwise.

Notice that the evaluation of the integral above must take into account that $\mathbf{x} + s\,(\cos\omega, \sin\omega)'$ may move through all sets $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{C} = \{\mathbf{y} \in \mathbb{R}^2 : 0 \leq y_1 \leq z_1, y_2 > 0\}$ for $\omega \in (\omega_0, \omega_1)$. This means that the formula for $d(\mathbf{x} + s\,\mathbf{U})$ changes while ω moves from ω_0 to ω_1. As a consequence, the interval (ω_0, ω_1) must be divided into several subsets by determining the angles at which there are transitions between the sets $\mathcal{L}$, $\mathcal{C}$, and $\mathcal{R}$. In general, the transitions between $\mathcal{R}$ and $\mathcal{C}$ happen at $\omega_2 = \arcsin(h_2) \geq \omega_3 = \arccos(h_3)$ with $h_2 = h_3 = (z_1 - x_1)/||\mathbf{x} - \mathbf{z}||$ whereas the transitions between $\mathcal{L}$ and $\mathcal{C}$ occur at $\omega_4 = \arcsin(h_4) \geq \omega_5 = \arccos(h_5)$ with $h_4 = h_5 = -x_2/||\mathbf{x} - \mathbf{z}||$. The angles ω_4 and ω_5 do exist if and only if

$x_2^2 \geq 2\, x_1\, z_1 - z_1^2$. If so then geometrical considerations lead to the relations $\omega_0 \leq \omega_2$ and $\omega_3 \leq \omega_5 \leq \omega_4 \leq \omega_2 \leq \omega_1$. Thus, there are four subcases depending on the value of ω_0. If ω_4 and ω_5 do not exist then only two cases must be considered.

If $\mathbf{x} + s\,\mathbf{U} \in \mathcal{C}$ then $d(\mathbf{x} + s\,\mathbf{U})/d(\mathbf{x}) = x_2/||\mathbf{x} - \mathbf{z}|| + \sin\omega$ which can be easily integrated. The situation changes if $\mathbf{x} + s\,\mathbf{U} \in \mathcal{L}$ or $\mathcal{R}$. In both cases one obtains an expression of the type

$$\frac{d(\mathbf{x} + s\,\mathbf{U})}{d(\mathbf{x})} = \sqrt{c} \cdot \sqrt{1 + a\, \cos\omega + b\, \sin\omega} \tag{9}$$

with $v^2 = a^2 + b^2 \leq 1$. The integral of this expression cannot be expressed in terms of elementary functions in general. In fact, using the identity $\cos\varphi\, \cos\omega + \sin\varphi\, \sin\omega = 1 - 2\, \sin^2((\varphi - \omega)/2)$, where $\varphi = \arccos(a/v)$ if $b \geq 0$ and $\varphi = \arcsin(a/v)$ otherwise, one gets

$$\begin{aligned} \int_{\theta_0}^{\theta_1} (1 + a\, \cos\omega + b\, \sin\omega)^{1/2}\, d\omega \quad &= \\ 2\,(1+v)^{1/2} \left[E\left(\frac{\varphi - \theta_0}{2}, m\right) - E\left(\frac{\varphi - \theta_1}{2}, m\right) \right] & \end{aligned}$$

where $m = 2\, v/(1+v) \leq 1$ and

$$E(\beta, m) = \int_0^{\beta} (1 - m\, sin^2\alpha)^{1/2}\, d\alpha$$

denotes the incomplete elliptic integral of the second kind [8]. If $\mathbf{x} + s\,\mathbf{U} \in \mathcal{L}$ then $(a, b, c) =$

$$\left(\frac{2\, x_2\, ||\mathbf{x} - \mathbf{z}||}{||\mathbf{x}||^2 + ||\mathbf{x} - \mathbf{z}||^2}, \frac{2\, x_2\, ||\mathbf{x} - \mathbf{z}||}{||\mathbf{x}||^2 + ||\mathbf{x} - \mathbf{z}||^2}, 1 + \frac{||\mathbf{x}||^2}{||\mathbf{x} - \mathbf{z}||^2} \right)$$

and if $\mathbf{x} + s\,\mathbf{U} \in \mathcal{R}$ then

$$(a, b, c) = \left(\frac{x_1 - z_1}{||\mathbf{x} - \mathbf{z}||}, \frac{x_2}{||\mathbf{x} - \mathbf{z}||}, 2 \right) .$$

Notice that the integral over the quotient $d(\mathbf{x} + s\,\mathbf{U})/d(\mathbf{x})$ is invariant with respect to scale, i.e., the replacement of $\mathbf{x}$ by $\xi\,\mathbf{x}$ and of $\mathbf{z}$ by $\xi\,\mathbf{z}$ with some $\xi > 0$ changes neither the parameters (a, b, c) in eqn. (9) nor the value of the angles ω_i. Therefore one may choose any $z_1 > 0$, say $z_1 = 100$, to calculate the values for $\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U})\,]/d(\mathbf{x})$ over the "critical" subset of $\mathcal{R}$ numerically. As can be seen from Fig. 5, the values are less than 1 and they rapidly approach unity in the vicinity of the point $(z_1, 0)'$. With some additional effort it may be shown analytically that the relation $\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U})\,] \leq d(\mathbf{x})$ is valid for all $\mathbf{x} \in \mathcal{R}$. This leads to the conclusion that inequality (8) does indeed hold true which in turn implies that the precondition of Theorem 1 is fulfilled for all $\mathbf{x} \in \mathcal{L} \cup \mathcal{R}$.

It remains to investigate the case $\mathbf{x} \in \mathcal{C}$ with $d(\mathbf{x}) = s = x_2 \geq 0$. Again, several subcases must be considered separately. Now let $\mathbf{p}$ and $\mathbf{q}$ denote the points of intersection of the two circles $\{\mathbf{y} \in \mathbb{R}^2 : ||\mathbf{y} - \mathbf{x}|| = s\}$ and $\{\mathbf{y} \in \mathbb{R}^2 : ||\mathbf{y}|| = ||\mathbf{x}||\}$. The interval (ω_0, ω_1) with the property

$$\omega \in (\omega_0, \omega_1) \iff f_0(\mathbf{x} + s\,\mathbf{U}) < f_0(\mathbf{x})$$

is specified by the angles $\omega_0 = \arccos(h_0)$ with $h_0 = (p_1 - x_1)/x_2$ if $x_1 > x_2/2$ otherwise $\omega_0 = \arcsin(h_0)$, and $\omega_1 = \arcsin(h_1)$ with $h_1 = (q_1 - x_1)/x_2$. The transitions between $\mathcal{C}$ and $\mathcal{L}$ occur at $\omega_2 = \arcsin(-x_1/x_2) \geq \omega_3 = \arccos(-x_1/x_2)$, whereas the transitions between $\mathcal{C}$ and $\mathcal{R}$ happen at $\omega_4 = \arcsin((z_1 - x_1)/x_2) \geq \omega_5 = \arccos((z_1 - x_1)/x_2)$. Depending on the actual location $\mathbf{x} \in \mathcal{C}$ it may happen that the pair of angles ω_2, ω_3, or the pair ω_4, ω_5, or even both pairs do not exist.

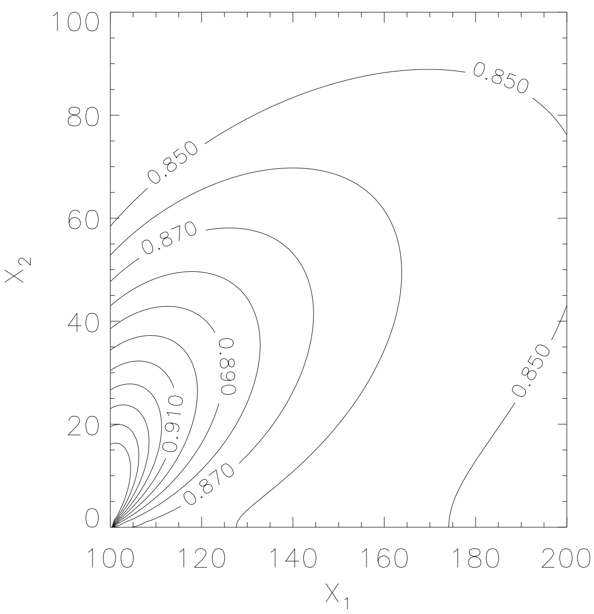

Fig. 5. Contour plot of $\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U})\,]/d(\mathbf{x})$ for $\mathbf{x} \in \mathcal{R}$.

If $\mathbf{x} + s\,\mathbf{U} \in \mathcal{C}$ then $d(\mathbf{x} + s\,\mathbf{U})/d(\mathbf{x}) = 1 + \sin\omega$ which is easily integrated. The other two subcases for $\mathcal{L}$ and $\mathcal{R}$ lead again to expressions of the type given in eqn. (9). The constants (a, b, c) are given by

$$\left(\frac{2\, x_1\, x_2}{x_1^2 + 2\, x_2^2}, \frac{2\, x_2^2}{x_1^2 + 2\, x_2^2}, 2 + \frac{x_1^2}{x_2^2} \right)$$

in case of $\mathbf{x} + s\,\mathbf{U} \in \mathcal{L}$ and by

$$\left(\frac{2\,(x_1 - z_1)\, x_2}{||\mathbf{x} - \mathbf{z}||^2 + x_2^2}, \frac{2\, x_2^2}{||\mathbf{x} - \mathbf{z}||^2 + x_2^2}, 1 + \frac{||\mathbf{x} - \mathbf{z}||^2}{x_2^2} \right)$$

in case of $\mathbf{x} + s\,\mathbf{U} \in \mathcal{R}$.

Figure 6 shows the values of $\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U})\,]/d(\mathbf{x})$ for $\mathbf{x} \in \mathcal{C}$. Evidently, the value is less than 1 unless $\mathbf{x}$ approaches the Pareto set. Therefore the region defined by the triangle $\Delta = \{\mathbf{x} \in \mathbb{R}^2 : 0 < x_2 \leq x_1 < z_1 - x_2\}$ deserves a closer inspection. Notice that $\mathbf{x} \in \Delta$ implies $\mathbf{x} + s\,\mathbf{U} \in \mathcal{C}$ for all $\omega \in (\omega_0, \omega_1)$. Consequently,

$$\begin{aligned} \frac{\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U})\,]}{d(\mathbf{x})} \quad &= \quad \int_{\omega_0}^{\omega_1} \frac{1 + \sin\omega}{2\,\pi}\, d\omega + \left(1 - \frac{\omega_1 - \omega_0}{2\,\pi}\right) \\ &= \quad 1 - \frac{\cos\omega_1 - \cos\omega_0}{2\,\pi} \\ &= \quad 1 - \frac{1}{2\,\pi} \left(\frac{q_1 - x_1}{x_2} - \frac{p_1 - x_1}{x_2} \right) \end{aligned}$$

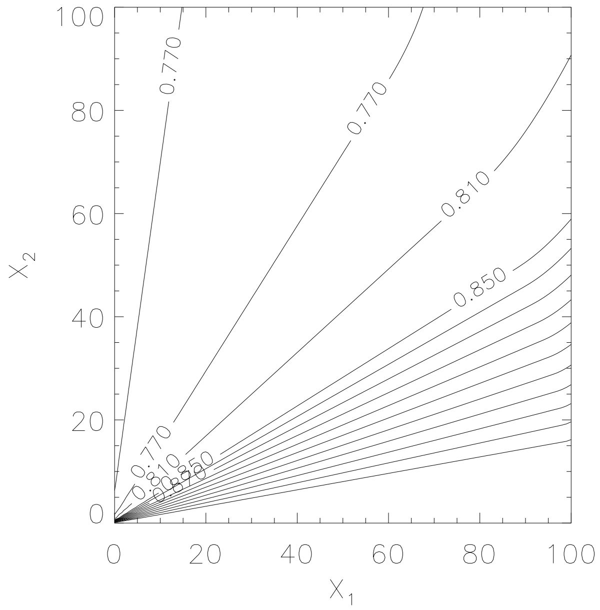

Fig. 6. Contour plot of $\mathsf{E}_0[\, d(\mathbf{x} + s\,\mathbf{U})\,]/d(\mathbf{x})$ for $\mathbf{x} \in \mathcal{C}$.

$$
\begin{aligned}
&= 1 - \frac{1}{2\,\pi} \cdot \frac{q_1 - p_1}{x_2} \\
&= 1 - x_2 \cdot \frac{(4\,x_1^2 + 3\,x_2^2)^{1/2}}{2\,\pi\,||\mathbf{x}||^2} \,. \qquad (10)
\end{aligned}
$$

For every $x_1 > 0$ the nonnegative rightmost term in eqn. (10) vanishes if and only if $x_2 = 0$. Since $d(\mathbf{x}) = x_2 > 0$ for $\mathbf{x} \in \Delta$ and owing to the symmetry property (6) one obtains

$$\mathsf{E}[\, D_{k+1} \,|\, \mathbf{X}_k = \mathbf{x} \,] = d(\mathbf{x}) - \gamma(d(\mathbf{x}))$$

with

$$\gamma(x_2) = \frac{x_2^2}{4\,\pi} \left(\frac{(4\,x_1^2 + 3\,x_2^2)^{1/2}}{x_1^2 + x_2^2} + \frac{(4\,(x_1 - z_1)^2 + 3\,x_2^2)^{1/2}}{(x_1 - z_1)^2 + x_2^2} \right)$$

provided that $\mathbf{x} \in \Delta$. Thus, the precondition of Theorem 1 is also fulfilled for $\mathbf{x} \in \Delta$. Summing up:

Theorem 2: Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be the sequence of parents generated by the multi–objective $(1+1)$–EA and let the objective functions be specified by eqn. (1). Then the $(1+1)$–EA converges with probability one to the Pareto set $\mathcal{X}^*$, i.e., $d(\mathbf{X}_k, \mathcal{X}^*) \to 0$ with probability one as $k \to \infty$. □

IV. Conclusions

It was shown that the multi–objective $(1+1)$–EA may converge with probability 1 to the Pareto set if the step size is proportional to the distance to the Pareto set. Although the proof itself is admittedly laborious—and only valid for a specific problem—it is an instructive example already disclosing the main difficulty which is inherent to this type of problems: There exists a subset of the search space in which the probability of accepting a point with smaller distance to the Pareto set decreases to $1/4$ whereas the probability of accepting a point with larger distance increases to $1/4$ as the parent moves closer and closer to the Pareto set. This fact makes the analysis not only complicated but it is also the reason why the mean rate of convergence to the Pareto set is subexponential in general.

Nevertheless, it is still an interesting question whether a similar result still holds in a more general situation or not. From a practical point of view it is necessary to find a method that realizes the step size rule without exploiting the knowledge about the location of the optimal solutions of both objective functions. In case of the special problem considered here, it is sufficient to know the gradients since

$$d(\mathbf{x}) = \frac{1}{2} \min_{0 \le \chi \le 1} \{\, ||\,(1-\chi)\,\nabla f_0(\mathbf{x}) + \chi\,\nabla f_1(\mathbf{x})\,||\,\} \,.$$

Notice that it is easy to determine the optimal χ analytically. If the objective functions are convex then it may be conjectured that $d(\mathbf{x})$ can be bounded from below and above by functions of the gradients. Provided that these bounds are sufficiently tight it may happen that the specialized result presented here is transferable to more general situations.

References

[1] C. M. Fonseca and P. J. Fleming. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16, 1995.

[2] H. Tamaki, H. Kita, and S. Kobayashi. Multi–objective optimization by genetic algorithms: a review. In *Proceedings of the 3rd IEEE International Conference on Evolutionary Computation*, pages 517–522. IEEE Press, Piscataway (NJ), 1996.

[3] J. Horn. Multicriterion decision making. In T. Bäck, D. B. Fogel, and Z. Michalewicz, editors, *Handbook of Evolutionary Computation*, pages F1.9:1–15. IOP Publishing and Oxford University Press, New York and Bristol (UK), 1997.

[4] G. Rudolph. Evolutionary search for minimal elements in partially ordered finite sets. In *Proceedings of the 7th Annual Conference on Evolutionary Programming (EP'98)*. Springer, Berlin and Heidelberg, 1998.

[5] F. Kursawe. A variant of evolution strategies for vector optimization. In H.-P. Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*, pages 193–197. Springer, Berlin and Heidelberg, 1991.

[6] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Kovač, Hamburg, 1997.

[7] R. S. Bucy and P. D. Joseph. *Filtering for Stochastic Processes with Applications to Guidance*. Interscience Publishers, New York, 1968.

[8] L. M. Milne-Thomson. Elliptic integrals. In M. Abramowitz and I. A. Stegun, editors, *Handbook of Mathematical Functions*, pages 587–626. Dover Publications, New York, 1965.